

Problem set 6

Selección de características Aprendizaje Automático

1. P1

Considere dos distribuciones unidimensionales con medias $\mu_1 = 2$ y $\mu_2 = 0$ y varianzas $\sigma_1^2 = \sigma_2^2 = 1$, respectivamente.

1. Genere 200 puntos de cada distribución.
2. Grafique sus histogramas respectivos.
3. Calcule y grafique el area bajo la curva de operación

Repita el experimento previo con a). $\mu_1 = \mu_2 = 0$ y b). $\mu_1 = 5$ y $\mu_2 = 0$. Comente sus resultados.

2. P2

Tenemos dos clases y asumimos que nuestras características son independientes y normalmente distribuidas en ellas. Las clases son modeladas con distribuciones gaussianas con medias $\mu_1 = [3 \ 3]'$ y $\mu_2 = [2,3 \ 2,3]'$ y matrices de covarianza $\Sigma_1 = 0,2I$ y $\Sigma_2 = 1,9I$. Calcule la divergencia entre las clases, la distancia de Bhattacharyya, la cota de Chernoff y grafique los datos.

Repita el experimento considerando los siguientes casos:

1. $\mu_1 = [3 \ 3]'$ y $\Sigma_1 = 0,2I$; $\mu_2 = [2,3 \ 2,3]'$ y $\Sigma_2 = 0,2I$.
2. $\mu_1 = [1 \ 1]'$ y $\Sigma_1 = 0,2I$; $\mu_2 = [4 \ 4]'$ y $\Sigma_2 = 0,2I$.
3. $\mu_1 = [1 \ 1]'$ y $\Sigma_1 = 0,2I$; $\mu_2 = [4 \ 4]'$ y $\Sigma_2 = 1,9I$

3. P3

En este problema, la medida J_3 se usará para escoger las mejores características. Consideraremos originalmente cuatro (media, desviación estándar, skewness y kurtosis). Tenemos diez valores de cada característica para cada clase.

Cargue los archivos *cirrhoticLiver.dat* y *fattyLiver.dat*.

1. Normalice los valores de cada característica para que tengan media cero y varianza unitaria.
2. Seleccione tres de las cuatro características que resulten con el mejor valor J_3 .
3. Grafique los datos en el espacio tridimensional de la mejor combinación de características.